

WHAT IS CLAIMED IS:

1. A system of cache residence prediction in a computer system, comprising:
a plurality of caches, comprising a first cache and one or more other caches; and
a prediction mechanism predicting that data requested by the first cache of a cache
5 miss can be found in at least one of the one or more other caches, if an address of the
cache miss matches an address tag of a cache line in the first cache and the cache line in
the first cache is in an invalid state.

2. The system of claim 1, wherein the prediction mechanism further
10 comprises a prediction table that records one or more addresses of one or more recently
replaced invalid cache lines.

3. The system of claim 1, wherein the prediction mechanism further
comprises a prediction table that records one or more addresses of one or more recently
15 replaced shared cache lines.

4. The system of claim 1, wherein the prediction mechanism further
comprises a prediction table that records prediction information for one or more
super-blocks, wherein each of the one or more super-blocks comprises one or more cache
20 lines, and wherein each entry in the prediction table includes a super-block address and a
valid vector that has a valid bit for each of the one or more cache lines.

5. The system of claim 1, wherein the prediction mechanism further
comprises a prediction table, wherein each entry in the prediction table corresponds to
25 one or more cache lines determined by a hash function.

6. A system of cache residence prediction in a computer system, comprising:
a plurality of caches, comprising a first cache and one or more other caches;
a memory and a memory controller operatively coupled to the memory; and

a prediction mechanism operatively coupled to the first cache, the prediction mechanism predicting whether data requested by the first cache of a cache miss can be supplied by at least one of the one or more other caches, producing a prediction result, and sending the prediction result to the memory controller;

5 wherein the memory controller uses the prediction result to determine if the memory is to be accessed immediately, or if the memory is not to be accessed until a corresponding cache snoop operation shows that the data requested by the first cache cannot be supplied by at least one of the one or more other caches.

10 7. The system of claim 6, wherein the prediction mechanism predicts that the data requested by the first cache of a cache miss can be supplied by at least one of the one or more other caches, if the address of the cache miss matches an address tag of a cache line and the cache line is in an invalid state.

15 8. The system of claim 6, wherein the prediction mechanism further comprises a prediction table that records one or more addresses of one or more recently replaced invalid cache lines.

20 9. The system of claim 6, wherein the prediction mechanism further comprises a prediction table that records one or more addresses of one or more recently replaced shared cache lines.

25 10. The system of claim 6, wherein the prediction mechanism further comprises a prediction table that records prediction information for one or more super-blocks, wherein each of the one or more super-blocks comprises one or more cache lines, and wherein each entry in the prediction table includes a super-block address and a valid vector that has a valid bit for each of the one or more cache lines.

11. The system of claim 6, wherein the prediction mechanism further comprises a prediction table, wherein each entry in the prediction table corresponds to one or more cache lines determined by a hash function.

5 12. The system of claim 6, wherein the memory can be a cache and the memory controller can be a cache controller of the cache.

13. A system of memory residence prediction in a computer system, comprising:

10 a plurality of caches, comprising a first cache and one or more other caches;
a memory and a memory controller operatively coupled to the memory; and
a prediction mechanism predicting whether data requested by the first cache of a cache miss can be found in at least one of the one or more other caches, and producing a prediction result;

15 wherein the prediction mechanism comprises a prediction table that is updated accordingly when the prediction mechanism observes a cache operation that can affect whether data requested by the first cache of a future cache miss can be found in at least one of the one or more other caches.

20 14. The system of claim 13, wherein the prediction mechanism records an address in the prediction table, if the prediction mechanism observes a cache operation indicating that requested data by the first cache of a future cache miss on the address can be found in at least one of the one or more other caches.

25 15. The system of claim 13, wherein the prediction mechanism removes an address from the prediction table, if the memory controller observes a cache operation indicating that requested data by the first cache of a future cache miss on the address cannot be found in at least one of the one or more other caches.

16. The system of claim 13, wherein the memory controller uses the prediction result to determine if the memory is to be accessed immediately, or if the memory is not to be accessed until a corresponding cache snoop operation shows that the requested data cannot be supplied by at least one of the one or more other caches.

5

17. The system of claim 13, wherein the prediction table records prediction information for one or more super-blocks, wherein each of the one or more super-blocks contains one or more cache lines, and wherein each entry in the prediction table further comprises a valid vector that has a valid bit for each of the one or more cache lines.

10

18. The system of claim 13, wherein each table entry of the prediction table corresponds to one or more cache lines determined by a hash function.

15

19. The system of claim 13, wherein the memory can be a high-level cache and the memory controller can be a cache controller of the cache.

20. A system of memory residence prediction in a computer system, comprising:

20

a plurality of caches, comprising a first cache and one or more other caches;

a memory and a memory controller operatively coupled to the memory;

a first prediction mechanism operatively coupled to the first cache, the first prediction mechanism predicting whether data requested by the first cache of a cache miss can be supplied by at least one of the one or more other caches, producing a first prediction result, and sending the first prediction result to the memory controller; and

25

a second prediction mechanism operatively coupled to the memory, the second prediction mechanism predicting whether requested data by the first cache of a cache miss can be supplied by at least one of the one or more other caches, and producing a second prediction result;

wherein an overall prediction result is determined by the first prediction result and the second prediction result; and

wherein the memory controller uses the overall prediction result to determine if the memory is to be accessed immediately, or if the memory is not to be accessed until a
5 corresponding cache snoop operation shows that the data requested by the first cache cannot be supplied by at least one of the one of more other caches.

21. The system of claim 20, wherein the overall prediction result comprises predicting that data requested by the first cache can be supplied by at least one of the one
10 or more caches, if the first prediction result or the second prediction result predicts that the data requested by the first cache can be supplied by at least one of the one or more caches.

22. The system of claim 20, wherein the memory can be a high-level cache
15 and the memory controller can be a cache controller of the cache.

23. A method of cache residence prediction in a computer system, comprising the steps of:

receiving a cache miss of a first address by a first cache;
20 predicting that data requested of the cache miss by the first cache can be supplied by one or more other caches, if the first address matches an address tag of a cache line in the first cache and the cache line is in an invalid cache; and

producing a prediction result and sending the prediction result to a component that can use the prediction result for some appropriate optimization.

25

24. A machine-readable medium having instructions stored thereon for cache residence prediction in a computer system, comprising the steps of:

receiving a cache miss of a first address by a first cache;

predicting that data requested of the cache miss by the first cache can be supplied by one or more other caches, if the first address matches an address tag of a cache line in the first cache and the cache line is in an invalid cache; and

5 producing a prediction result and sending the prediction result to a component that can use the prediction result for some appropriate optimization.